Statistical Analysis: Vektron® 6913 Fleet Test

Statistical Design and Analysis of Vektron® Emissions Fleet Trial

Infineum USA L.P

2/22/2001

(A.0) Introduction

(A.1) Abstract

A 28-vehicle fleet test was run to verify and quantify the NOx emissions reductions achieved by Infineum's Vektron 6913 gasoline additive. The fleet composition and experimental design were finalized in collaborative discussions with US Environmental Protection Agency (EPA) Office of Transportation & Air Quality (OTAQ) and consultation / advice from several major US automotive manufacturers (OEMs). The test was conducted over a period of five months (Feb to Jul/2000) at Southwest Research Institute (SwRI). Southwest Research Institute is an independent, nonprofit, applied engineering and physical sciences research and development organization. Statistical analysis of the emissions data indicated a 10% average fleet reduction in NOx emissions without any negative impact on other criteria pollutants (CO, HC) or fuel economy. The details around items such as vehicle and fuel selection, level of additive are covered elsewhere and will not be covered here. The purpose of this report is to document the design and analysis methodologies used in the analysis.

(A.2) Summary

(A.2.1) Design

The experimental design was a modified two period crossover design, which included a 1,000 mile baseline run with reference fuel, and two 8,000 mile periods (Run 1 and Run 2) with either reference fuel, or a Vektron present fueling scheme. There were two Vektron present fueling schemes. In one case, Vektron was in each tank of fuel. In the other case fuels switched between a tank with reference fuel and one containing Vektron. In other words, every other tank contained Vektron. An a-priori success criterion of ($p<.10$) was established for statistical significance. With crossover designs Run 1 is a valid parallel test in itself and provides an unbiased estimate of an effect. A date was therefore set to analyze the Run 1 data.

(A.2.2) Results Run 1

On the date specified about 92% of the data was in. There were a few problems. One vehicle was replaced for mechanical problems early in the test. Two statistical outliers were found. The cause of one outlier was identified as a clogged EGR hose. The second outlier was determined to be the vehicle with the highest oil consumption (though it is not known if this was the cause). Both outliers had studentized residuals > 3.7 and the vehicles were dropped from the analysis.

The first finding was there was no debit from fuel switching between a reference and Vektron fuel. ***The second finding was that the effect of Vektron was statistically significant at ($p< .06$) and the effect was estimated to be 10 %.*** Consideration was given to stopping the trial at that time since the effect had been proven and could be estimated. The decision was made to complete the test.

(A.2.3) Results Run 2

Early Run 2 analysis seemed to indicate that there might be some carryover. This finding was believed to an artifact of incomplete data at that time. Additional data would probably cause the carryover to shrink or even disappear. Freeman's (1989) warns that additional data simply exacerbates the problem when carryover is present. Indeed, the carryover effect continued to be present and became more pronounced as the remaining data from Run 2 came in. The carryover effect was large and aliased with miles. It was decided that no reliable estimate could be made using both Run1 and Run 2 data. A decision was made to only estimate the Vektron present effect from Run 1.

(A. 3) Structure of this report

The report has seven sections:

A) Introduction,

B) Experimental design

C) Description of the Data

D) Brief discussion of the statistical methods used

E) Results and Analysis

F) References

G) Appendix.

As stated, the first analysis was done with Run 1 data. Most of the statistical analysis described in this report (Section E) describes analysis done on Run 1 data. The reason for examining only the Run 1 data was that when the Run2 was complete and the entire dataset was examined, a large carryover effect was observed which was aliased with mileage and treatment. This carryover effect made it impossible to use the Run 2 data in the estimation process (A discussion of the carryover can be found in section E.5.). However, the Run 1 data set is a stand alone parallel study and provides an unbiased estimate of the effect.

(B . 0) Experimental Design

(B. 1) Vehicles

The makeup of the vehicle types was a group agreed upon by EPA OTAQ in consultation with OEM experts. In this design vehicle type represents random blocks from the population of vehicle types in the car park. The table below lists these vehicle types.

(B.1.1) Table: Test Vehicles

Vehicle Type	Vehicle Type Code	Power Plant	Fuel System	Transmission	Model Year	Certification
Ford Explorer	**EX**	**4.0L V-6**	**MPFI**	**Automatic**	**'99**	**LEV LDT**
Chevrolet -1500	**GC**	**5.7L V-8**	**CSFI**	**Automatic**	**'99**	**LDT**
Honda Accord	**HA**	**2.3L I-4**	**MPFI**	**Automatic**	**'98**	**LEV LDV**
Ford F-150	**FF**	**4.6L V-8**	**MPFI**	**Automatic**	**'97**	**LDT**
Ford Escort	**FE**	**1.9L I-4**	**MPFI**	**Automatic**	**'96**	**LDV**
Dodge Caravan	**DC**	**3.3L V-6**	**MPFI**	**Automatic**	**'95**	**LDT**
GM Buick LeSabre / Olds 88 Royale	**GP**	**3.8L V-6**	**MPFI**	**Automatic**	**'94**	**LDV**

(B. 2) Design Matrix

Crossover designs have a long history of being effective and efficient methods for testing effects, as long as no large carryover is present. This design was selected in conjunction with EPA OTAQ and consultations with representative OEMs from a set of possible designs. The assumption was that even if carryover were present, it would be small. Also, the 8,000 miles would provide a sufficient washout period.

The design matrix had the following form:

(B.2.1) Table: Field Test Design

Veh #	Veh Code	Veh Type	Fuel Scheme Group (FSG)	Start of Test (SOT)	Run 1	Run2
		Random Block	**CF=Constant Fueling** **AF=Alternating Fueling**	1000 miles	8,000 miles	8,000 miles
1	EX1	EX	CF	Ref	Ref[1]	Vektron[2]
2	EX2	EX	CF	Ref	Vektron	Ref
3	EX3	EX	AF	Ref	Ref	Vektron
4	EX4	EX	AF	Ref	Vektron	Ref
:	. :	: :	: :	: :	: :	: :
28	GP4	GP	**AF**	Ref	Vektron	Ref

1. Ref - Reference fuel is in every tank during the run
2. Vektron is present during the run

The above table summarizes the design. One vehicle was replaced (FE3) with (FE5) early in testing because of mechanical problems (see 3.2.2).

A description of each of the columns of the above table can be found below.

<u>Column</u> <u>Description</u>

Veh #: The number of vehicles (1 to 28).

Veh Code: The code assigned by SwRI for each vehicle.

Veh Type: The vehicle type (e.g. Explorer, see table B.1.1) and is the random block in this design. There were 7 vehicles types, with 4 vehicles in each block.

FSG: The Fueling Scheme Group (FSG): Constant Fueling (CF), or Alternating Fueling (AF).

- <u>Constant fueling</u> (CF): In Run 1 or Run 2 when Vektron was present in each tank of fuel.

- Alternating fueling (AF): Vektron was present in every other tank of fuel. In other words, the fuels alternated between the reference fuel and the fuel with Vektron. This was done to simulate the effects of brand switching by the consumer.

- Example: For example, the first vehicle in the table (EX1) ran the Constant Fueling (CF) scheme, which meant every tank in Run 2 had Vektron in it. In contrast, the third vehicle (EX3) used the Alternating Fueling (AF) scheme. In Run 2 fuel tanks alternated between Vektron and the reference fuel.

SOT: Every vehicle ran the reference fuel (Ref) for 1,000 miles to establish a baseline.

Run 1: This was the first 8,000 miles of the crossover design. Vehicle (EX1) ran (SOT), then a reference (Ref) fuel for 8,000 miles.

Run 2: This was the second 8,000 miles of the crossover design. Vehicle (EX1) ran a Constant Fueling (CF) scheme with Vektron present in Run 2.

The above table illustrates only one complete block: the Explorer vehicle type. This basic block design was repeated for each of the 7 vehicle types. Vehicles were randomly assigned within a block. The exception to this was when it was difficult to obtain vehicles of a particular type. However, because the vehicles came in from different locations (in some cases different states), this was itself seen as a random process.

(C . 0) Data

(C. 1) Data QC Procedures

Quality control procedures were implemented to ensure data integrity. A data chain of custody between SwRI, Infineum, and groups within Infineum was established. A database was developed which included all the data on the emissions test printout. The database was programmed to simulate the calculations produced by the analytical equipment at SwRI to check for errors. Other checks looked for duplicates and typos. In addition, the entries in the database were checked manually against each data sheet received from SwRI. The database was also reviewed in detail with SwRI and against the final SwRI test program report.

(C. 2) Emissions Testing

Vehicles were run and emissions testing were done at Southwest Research Institute (SwRI). The FTP 75, USO6, and HFET emissions tests were done on these vehicles, though this paper will deal with the weighted FTP (wtdFTP) in the FTP75 test since it is the primary test of interest. Vehicles were tested after SOT, at the end of Run 1, and again at the end of Run 2.

(C. 2. 1)Auto/Oil Procedure and Means

Each vehicle at each test point was tested in duplicate over the FTP, HFET, and the US06 driving cycles. Regulated exhaust emissions (NOx, CO, and NMHC) were recorded for each test. The need for a third test sequence was based on test repeatability criteria established by Painter and Rutherford in SAE paper 920319 and the CRC Auto/Oil Protocol. A third FTP test was run if necessary as judged using the CRC Auto/Oil Protocol: (if higher emission value divided by the lower emission value is >1.33 for HC, >1.70 for CO, or >1.29 for NOx, run a third test).

There were two exceptions to the above protocol: one vehicle had to be restarted, and another was eventually dropped for mechanical problems. Test results for each fuel were averaged and the average value was used in the final statistics (a convention adopted elsewhere -- see Ragazzii, R., Nelson, K., Mobil Sources Program, Colorado Department of Public Health and Environment, 1999,*pg.21)*.

(D . 0) Methods

The purpose of this section is to introduce some of the methodologies used to obtain the results in section (E.). It is not meant to be comprehensive, but simply to provide some background to the reader. More extensive treatments for each of these subjects can be found elsewhere (Vonesh and Chinchilli, 1997, Pinheiro and Bates, 2000).

(D.1) Random Vs Fixed Effects

Factor effects can be either Random or Fixed. Fixed effects are those effects estimated by changing different levels of a factor or testing different known factors or treatments. The fixed effects included the presence of Vektron in the fuel, and a nested fueling scheme (FSG) factor. The random block in this study was vehicle type.

(D.2) Blocking

Blocking is a technique used to diminish the effects of variation among experimental units. Blocks are groups of homogeneous units and the treatments are randomly assigned to units within the blocks. Block effects are considered random because the blocks in the experiment are only a subset of the larger set of blocks over which inference about treatment is to be applied. Vehicle type is the blocking variable selected to represent the larger population of vehicle types (i.e. the car park).

(D.3) Simple Linear Models

There is general familiarity with the simple linear model. This model can be used to test for the effects of a factor that is changed in an experiment (e.g. treatment). The factor is modeled as:

A **treatment (α)**,
Around a mean **(μ),**
In the presence of **random error (*e*),**
With the following relationship to the **response (y):**

$$y_{i,j} = m + a_i + e_{i,j} \qquad \text{(D.3.1)}$$

$\varepsilon_i \sim N(0, \sigma^2 \mathbf{I})$

The model can be generalized to multiple levels or treatments, for example $(a_A \ldots. a_D)$. The error in this model is usually unknown and is estimated from the data to provide estimates for the standard error to perform statistical tests on the treatment. These models assume that errors are not correlated. They therefore may not handle within block correlation well.

(D.4) Linear Mixed-Effects Models

Mixed-Effects models are a flexible class of methods which can be used in experiments where there are fixed effects within random blocks. The mixed-effects methods of analysis (such as those employed by S-plus lme and SAS Proc Mixed) give good estimates even when there is correlation within blocks.

(D.4.1) Single Level of Grouping

The linear mixed-effects model for a single level as described by Laird and Ware (1982) gives the n_i-dimensional response vector $\boldsymbol{y_i}$ for the ith group as

$$y_i = \mathbf{X_i} b + \mathbf{Z_i} b_i + e_i \quad i = 1,\ldots M \qquad \text{(D.4.1.1)}$$

$b_i \sim N(0, \Sigma)$ $\quad \varepsilon_i \sim N(0, \sigma^2 \mathbf{I})$

Where

β= p-th dimensional vector of fixed *effects*,

b_i=q-dimensional vector of *random effects,*

$\mathbf{X}_i = (n_i\ x\ p)$ fixed effects regressor matrix

$\mathbf{Z}_i = (n_i \text{ x } q)$ random effects regressor matrix, and

ε_i is the n_i – dimensional within group error vector.

The assumption $Var(\varepsilon)=\sigma^2 \mathbf{I}$ can be relaxed with S-plus Linear Mixed Effect (**lme**) Procedure (See Pinheiro and Bates, 2000) because of its ability to model non constant variances or special within-group correlation structures.

(D.4. 2) Multi Level of Grouping

These effects can be nested. In this study the FSG is nested within the random block (vehicle type). Likelihood and restricted likelihood functions are calculated by S-plus's l**me** procedure using the same techniques used for the single grouping case in (D.4. 1). The reader is referred to Pinheiro and Bates, 2000.

(D.4.3) Likelihood Ratio Test

The necessity to include terms in the model should always be checked. When using linear mixed effects models a likelihood ratio test provides a formal model selection and a goodness of fit test using normal assumptions for nested models to test for the need for terms in the model.

Imagine comparing two models, a simple model and one containing an additional term.

$$\text{Model 1: } y_i = [X_{i1}, X_{i2}]\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + Z_i b_i + e_i \qquad \text{(D.4.3.1)}$$

Versus

$$\text{Model 2: } y_i = [X_{i1}]\left[b_1\right] + Z_i b_i + e_i \qquad \text{(D.4.3.2)}$$

The likelihood ratio test is calculated by taking the difference in 2 times the log-likelihoods from the two models.

$$\text{Likelihood-Ratio: } 2L_1(\widehat{b}_1\widehat{b}_2, \widehat{q}) - 2L_2(\widehat{b}_1, \widehat{q}) \sim c_{s_2}^{\;2} \qquad \text{(D.4.3.3)}$$

For additional discussion see Vonesh and Chinchilli, 1997.

The S-plus **anova** procedure provides a method for comparing these models and the impact of additional variables. In the results section this is used to test the effects of removing the FSG term.

(D.4.4) SAS Proc Mixed vs. S-Plus lme

Most of the analysis done in this work was done using S-Plus. However, the final models were tested using both S-Plus and SAS Proc Mixed. The results calculated by two software procedures were nearly identical.

(E . 0) Results

(E.1) Data Transformations

There is significant literature in the emissions area using either percent difference from start of test (SOT = start of Run 1) or difference in natural log (ln) difference from start of test. Both were used in this study. These are defined in (E.1.1, and E.1.2).

Percent difference from SOT:

$$dPercDiff_{i,j} = (y_{i,j}-SOT_{i,j}) / SOT_{i,j} \qquad (E.1.1)$$

Difference in natural log (ln) from SOT:

$$dlnDiff_{i,j} = \ln(y_{i,j}) - \ln(SOT_{i,j}) \qquad (E.1.2)$$

Where y = weighted FTP-75
i = 1 to 3 (SOT, Run 1, Run 2)
j = 1 to 28 (28 vehicles)

(E.2) Outliers and Mechanical Problems

(E.2.1) Outliers

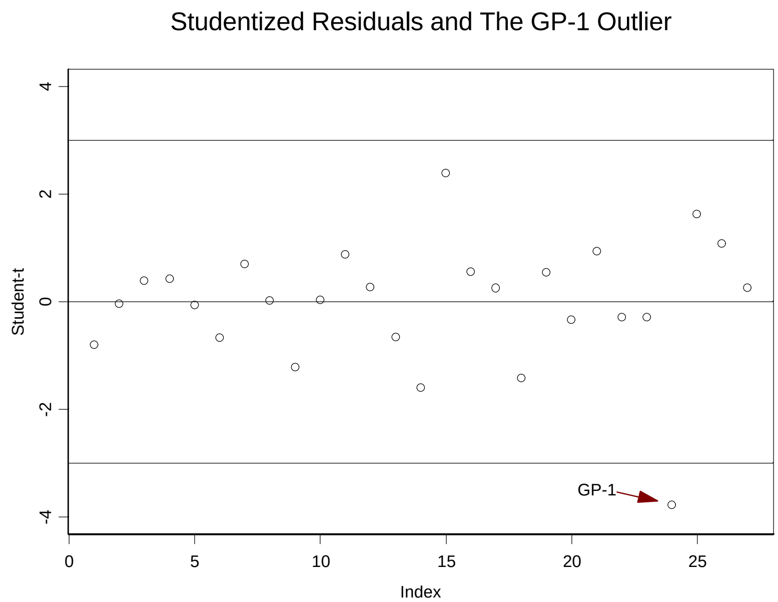
During the design stage of this program, the Infineum project team and the EPA OTAQ technical team set up an a-priori decision that statistical outliers could be eliminated from data analysis if an attempt was made to discover the cause.

After the completion of Run 1 an interim data analysis was done. Mixed-Effects models can be sensitive to outliers. It is therefore necessary to screen for these outliers before fitting the Mixed-Effects model. Because of its diagnostics ability, a simple general linear model (glm) was fit with vehtype and additive both fit as fixed effects. In linear systems Studentized residuals should be included as a routine part of the diagnostic phase of any analysis (Beckman,R.J. and Cook,R.D. 1983). Models fit to the Run 1 data and the residuals from the model were plotted and tested as outliers. Vehicles FF4 and GP-1 were identified as outliers.

- FF-4 was discovered during the first analysis of Run 1 (before GP-1 Run 1 was completed). When the response PercDiff is modeled against the variables (vehtype and additive present) using (SAS) PROC GLM, the absolute value of the RSTUDENT (or studentized residual if the point is removed) statistic is *greater than 4.* In this case, an investigation led to a clear physical cause for this outlier (see 3.2.2). The data was dropped.

- Run 1 data from vehicle GP-1 was also dropped as a statistical outlier. An explanation for this outlier may be high oil consumption since this vehicle had the highest oil consumption in the qualification phase. Again, fitting PercDiff with PROC GLM the absolute value of the RSTUDENT statistic is calculated. *For GP-1 the t-statistic is greater than 3.78.* The graph below is a plot of the Studentized Residuals by observation number and demonstrates the aberrant nature of GP-1.

(3.2.1.1) Figure: Outlier Test and Studentized Residuals



(E.2.2) Identified Mechanical Problems

Several mechanical difficulties were experienced during this test program:

- FE-3: this Ford Escort experienced problems with the oxygen sensor early in the test. The oxygen sensor was replaced and the test restarted; however, further problems with the catalyst system followed. This vehicle was replaced with another Ford Escort, which was coded as FE-5 to clearly keep the vehicle histories separate.

- FF-4: higher than expected NOx emissions results were generated after Run 1 and Run 2 emissions testing on this Ford F-150. A preliminary analysis of the Run 1 data had indicated that this vehicle could be an outlier. The vehicle's exhaust gas recirculation (EGR) valve was inspected after Run 2 and deposits were found to be completely plugging one of the two EGR ports. The valve was replaced and a second set of Run 2 emissions tests were performed on FF-4 resulting in expected levels of emissions. Because this change broke the test protocol, the FF-4 Run 2 test data could not be used in the data analysis. This plugged EGR ports could also explain why the Run 1 data was a statistical outlier. It was too late in the test program to replace this vehicle, so it was removed from the analysis.

Overall, two vehicles out of the 28 test vehicles were removed from data analysis as statistical outliers. One vehicle was replaced during the test because of mechanical difficulties and the new vehicle was included in the data analysis.

(E.3) FSG: Alternating (AF) vs Constant (CF) Fueling

Each vehicle was randomly assigned to one of two fueling schemes (FSG) within the vehicle type block. The description is given in (B.2.1).

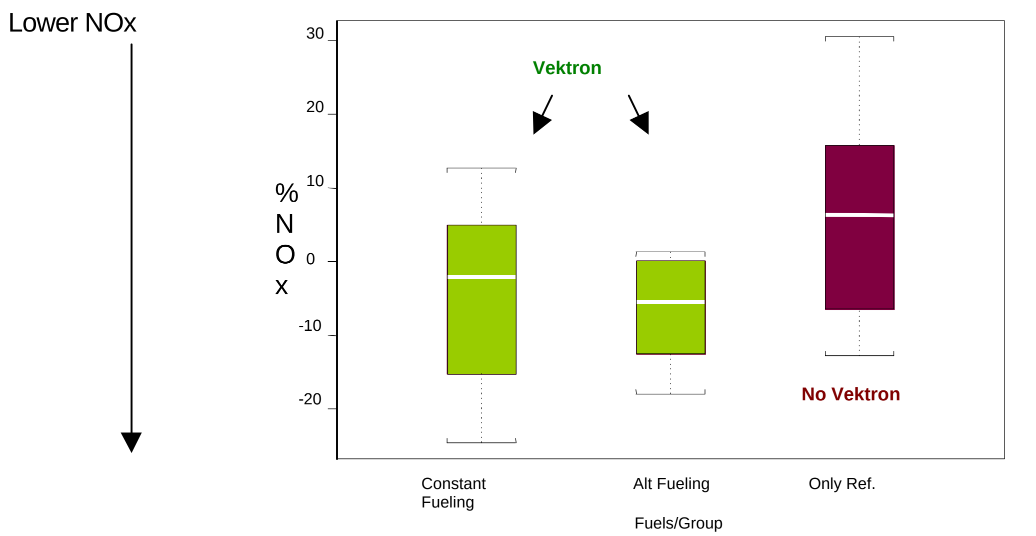
(E.3.1) Simple graphical inspection

A model was fit to the response PercDiff using only the random effects blocking term (VehType). This model did not contain the Fueling Scheme (FSG) or Vektron present fixed effects terms. The model was simply:

$$y_i = Zb_i + e_i \qquad \text{(E.3.1.1)}$$

By removing the effects of VehType and by plotting the model residuals vs the type of fueling run during the Run 1, the effects of FSG and Vektron present are isolated and observed for the first 8, 000 miles.

(E.3.1.1) Figure: VehType Model Residuals vs Fuels/Group



It is apparent that the AF (which was expected to be between the Reference and CF) has at least as large an effect as the Constant Fueling (CF) case. This simple analysis suggests that there is no debit associated with fuel switching and that is reasonable to collapse the FSG group into the single term for Vektron Present.

(E.3.2) Likelihood Ratio Test: FSG Term in the Model

A more formal way of determining if there is a need for the FSG term in the model is to use a Likelihood Ratio Test. This test is often used when models are fit using a maximum likelihood (ML) method. Although the restricted maximum likelihood method (REML) gives the best unbiased estimate, the Likelihood Ratio Tests may not be appropriate for testing in some cases for models fit with REML. However, the likelihood ratio test can be used to compare models if both models were fit using REML (see page 83, Pinheiro and Bates, 2000). The models presented here were all fit using REML. These models were also fit by ML and the Likelihood Ratio Test was run with similar conclusions for the FSG term (see Appendix G.3).

Two groups of models were fit, one group for each transformed response from (E.1):

1) Models with the <u>FSG term in the model</u>:

Model (1a) for Response=dlnDiff from E.1,
Model (1b) for Response=PercDiff

2) Models with <u>FSG dropped from the model:</u>

Model (2a) for Response=dlnDiff,
Model (2b) for Response=PercDiff.

First, the models were fit. The ANOVA tables can be found in the next section.

(E.3.2.1) Model(1a): Resp= dlnDiff, with FSG term

The S-plus **lme** code is:

Mod1a_lme (dlnDIff~AdditiveRun1, **random=~1|vehtype/FSG**)

Model (1a) models the effects of Vektron Present (AdditiveRun1) with FSG in the random block VehType on log difference (dlnDiff). This model contains all terms and is the complete model. The analysis of variance (ANOVA) for this model is the following:

(E.3.2.1.1) Table: ANOVA of model (2a) with *FSG*

	NumDF	denDF	F-Value	p-value
(Intercept)	1	11	5.445504	0.0396
AdditiveRun1	1	11	6.101462	0.0311

This model appears to give a very good fit, and a significant Vektron present effect (p<. 032). The simpler model below is a fit of the data with the FSG term removed.

(E.3.2.2) Model (2a): Resp= dlnDiff, without FSG Term

The S-plus **lme** code is:

Mod2a_lme (dlnDiff~ AdditiveRun1, random=~1|vehtype)

(E.3.2.2.1) Table: ANOVA of simple model (1a) *without FSG*

	numDF	DenDF	F-value	p-value
(Intercept)	1	18	4.888377	0.0402
AdditiveRun1	1	18	4.102633	0.0579

The significance level dropped from (p<.032) to (p<. 06). However, the effect of removing the term needs to be tested using a Likelihood Ratio Test.

(E.3.2.3) Likelihood Ratio Test: models (1a and 2a)

Running the S-plus **anova** procedure produces the following result.

(E.3.2.3.1) Table: Likelihood Ratio Test: Models (1a, 2a)

	Model	df	AIC	BIC
Mod1a	1	5	-15.11304	-9.22277
Mod2a	2	4	-14.99587	-10.28366
	LogLik	Test	L.Ratio	p-value
Mod1a	12.55652			
Mod2a	11.49794	1 vs. 2	2.117168	0.1457

The last line of the table above has the results of the Likelihood Ratio Test. The small difference in the Likelihood Ratio (L.Ratio) and the non-significant p-value (p<. 15) indicates the FSG term can be removed. This confirms the simple graphical analysis (E.3.1.1). It is therefore reasonable to collapse to a single fixed effect term for Vektron Present.

(E.3.2.1.1) Model (1b): Resp=PercDiff, with FSG term

The analysis was repeated for the PercDiff (see E.2.1) response: (mod1a, and mod1b).

(E.3.2.1.1) Table: ANOVA for model (2b*) with FSG*

.	NumDF	denDF	F-value	p-value
(Intercept)	1	11	5.898497	0.0335
AdditiveRun1	1	11	6.480273	0.0272

(E.3.2.1.1) Model (2b): Resp-=PercDiff, FSG term dropped

(E.3.2.1.1) Table: ANOVA Simple model (1b) *without FSG*

.	NumDF	DenDF	F-value	p-value
(Intercept)	1	18.0000	5.3489	0.0328
AdditiveRun1	1	18.0000	4.2665	0.0536

(E.3.2.3) Likelihood Ratio Test: Models (1b and 2b)

(E.3.2.3.1) Table: Likelihood Ratio Test models (1b and 2b)

	Model	Df	AIC	BIC
Mod1b	1	5	-9.498291	-3.608021
Mod2b	2	4	-9.325839	-4.613624
	LogLik	Test	L.Ratio	p-value
Mod1b	9.749145			
Mod2b	8.662920	1 vs. 2	2.172451	0.1405

The analysis indicates that it is reasonable to collapse to a single fixed effect term for Vektron Present.

(E.3.2.4) Test for FSG as a fixed effect

For completeness, FSG is pulled out of the nested position within vehicle type and tested as a simple fixed effect term. The model is refit using S-plus.

S-code: modfixed.lme_lme(PercDiff~AdditiveRun1+ FSG, random=~1|vehtype)

ANOVA Table Tests for FSG as Fixed Effect (Not Significant)

	NumDF	DenDF	F-value	p-value
(Intercept)	1	17.00	5.5675	0.0305
AdditiveRun1	1	17.00	4.3274	0.0529
FSG	1	17.00	1.5448	0.2308

The FSG term is not significant ($p< .23$). The simple graphical analysis and the Likelihood Ratio Tests are confirmed. There is therefore plenty of evidence to drop the FSG term and only use the Additive Present Term.

(E.4) Estimating Additive Effect on Nox

The next step was to estimate the effect of Vektron present from Run 1 data.

(E.4.1) Simple analysis of Run 1 Data Means

A very simple and rough analysis of means is given in the table below. The analysis begins by calculating mean difference from the start of test (SOT) to the end of Run 1 for each vehicle. The mean difference is calculated for those vehicles in a vehicle type block run when Vektron was not present (i.e. Ref fuel) during Run 1. These means are column 2 (Untreated). Next the mean difference is calculated for those vehicles that had Vektron present during Run 1. These means are placed in column three (Treated). By subtracting the Treated group (column 3) from the Untreated group (cool 2), the Treat Effect (i.e. effect of Vektron present) is calculated. An overall average Treat effect is calculated by averaging down the vehicle types**.** *The number estimated was 12.04 % effect from treat.*

Vehicle Type	Mean PercDiff (Untreated)	Mean PercDIff (Treated)	Treat Diff = PercDiff(Treated)-PercDiff(Untreated)
Ford Explorer '99	3.2 %	2.2 %	-1.0 %
Chevrolet–1500 '99	11.3	-4.5	-15.8
Honda Accord '98	15.1	12.9	-2.2
Ford F-150 '97	22.8	-13.6	-36.4
Ford Escort '96	38.5	16.1	-22.4
Dodge Caravan '95	-6.1	-3.9	2.3
GM Buick LeSabre / Olds 88 Royale '94	35.5	26.7	-8.8
		Overall Mean	-12.04 %

Although a more precise estimation method is needed and will be discussed in the next section (E.4.2), the above analysis should serve as a touchstone for any more complicated analysis. If an analysis produces an estimate very far from the one obtained here it should be suspect. One final observation is that 6 out of the 7 vehicle types see an additive effect. This is an unlikely event by chance.

(E.4.2) Estimation Treat Effect on NOx Using Mixed-Effects Modeling

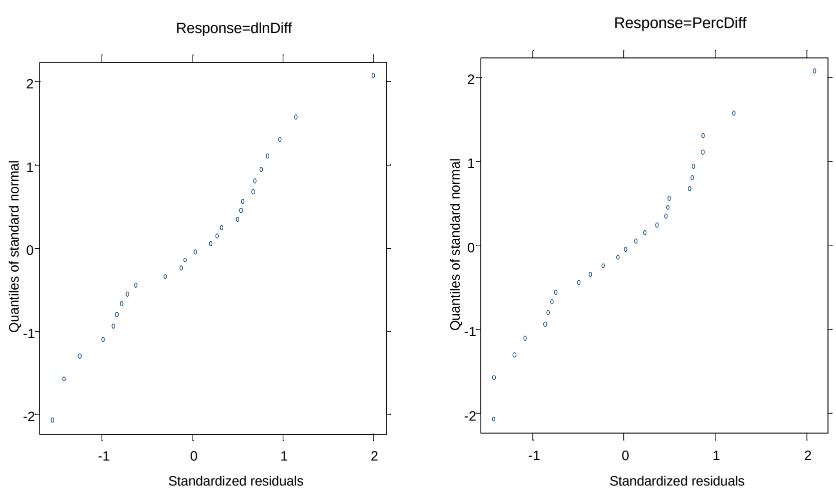
The assumptions of normality are checked in this section for the Mixed-Effects Model. In addition an estimate of the effect of Vektron in Run1 is obtained. The models described here are models (2a and 2b) from section E.3 where

Fixed Effect: AdditiveRun1
Random Block Effect: VehType.

(E.4.2.1) Normality Assumption

In (D.4.1.1) normality $\varepsilon_i \sim N(0, \sigma^2 \mathbf{I})$ for the errors was seen as an assumption for Mixed-Effects modeling. These errors can be checked by examining the model residuals. A normal probability plot provides a method for examining these residuals. Normal residuals are expected to fall along a straight line.

(E.4.2.1.1) Figure: Normal Probability Plot of Residuals



These plots indicate that the normality assumption is reasonable. Normality is tested more formally by performing a Kolmogorov-Smirnov test. The table below summarizes this analysis.

(E.4.2.1.1) Table: Kolmogorov-Smirnov Test for Normality

Model	p-value
Mod2a: dlnDiff	p> .5
Mod2b: PercDiff	p> .5

The large p-values in the above table confirms the assumption of normality.

(E.4.2.2) Normality Assumption for blocks (VehType)

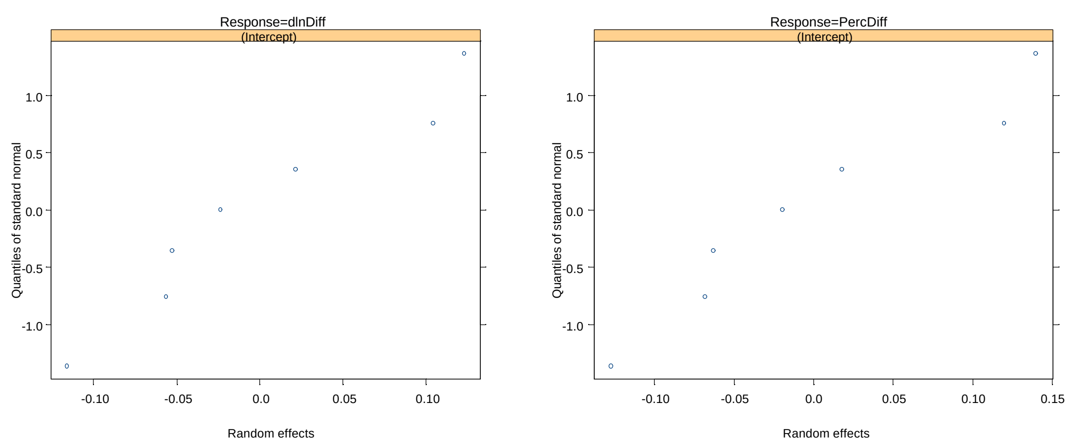
In addition to the assumption of normally distributed errors, the blocks are also assumed to be normal $b_i \sim N(0, \Sigma)$, i.e. that the block or random effects are normal. The following table summarizes the random block effects of vehicle type estimated with the Mixed-Effects models.

(E.4.2.2.1) Table: Random Effects from models (2a and 2b)

Block	Response=dlnDiff	Response=PercDiff
DC	-0.12	-0.13
EX	-0.06	-0.07
FE	0.10	0.12
FF	-0.02	-0.02
GC	-0.05	-0.06
GP	0.12	0.14
HA	0.02	0.02

Although there are not many blocks, they can still be examined using normal probability plots.

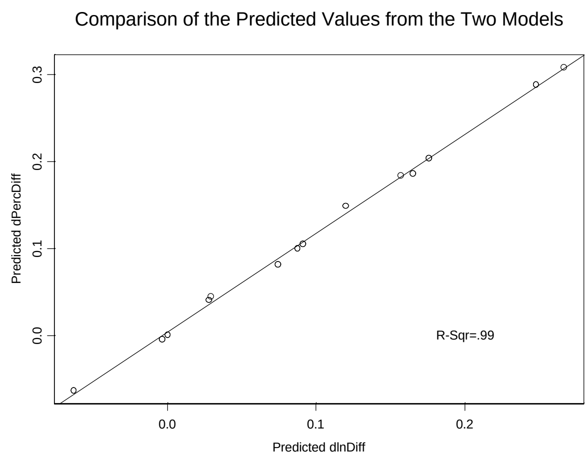
(E.4.2.2.1) Figure: Normal Prob Plot of Random Effects



Normality of the random effect (b) is a reasonable assumption.

(E.4.2.3) Comparison between PercDiff and dlnDiff

In (E.1), the two most frequently used transformations for NOx were described. These were the natural log of the Nox (dlnDiff), and the percent difference (PercDiff). The analysis indicated that the conclusions are the same using either response. The plot below shows that they are highly correlated with each other (as would be expected).



(E.4.2.4) Estimate of Vektron Effect on NOx

The expected effect of Vektron Present is estimated directly by fitting the model (2b).

Fixed effects: PercDiff1 ~ AdditiveRun1

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.1683203	0.05622366	18	2.993763	0.0078
AdditiveRun1	-0.1042360	0.05046390	18	-2.065556	0.0536

The last line of the coefficients table gives the estimate (Effect=-. 1042). The 90% confidence intervals are:

Table: Approximate 90 % confidence intervals for the fixed effects:

.	Lower	Est.	Upper
(Intercept)	0.0708249	0.1683203	0.26581571
AdditiveRun1	-0.1917436	-0.1042360	-0.01672839

Multiplying the estimate (-. 1042) by 100 expresses the result in percentage.

Effect of Vektron in the Fuel= -10.42 %. Rounding to the nearest unit gives -10 %.

This estimate is not very far from the simple analysis of means in (E.4.1). This is the identical estimate obtained when running the SAS software procedure PROC MIXED (see Appendix G.2).

As stated earlier there was a large carryover effect observed in the data when both Run 1 and Run 2 data were examined. The purpose of the next section is to discuss the nature of this carryover and why the data from Run 2 could not be used in making these estimates.

(E.5) Run 2 Data and Carryover

The main disadvantage of crossover designs is that carryover effects may be aliased with treatment effects. The assumption during the design stage was that carryover was not going to be a concern. Carryover is defined as the effect of treatment from the first time period on the current time period. Carryover is typically observed when the order of treatment by period leads to different effects on the response. Carryover effects cannot be estimated directly with a two period crossover design unless you make the assumption that carryover from the two treatments (A to B or B to A) are identical and cancel each other out. This assumption cannot be made in this study.

(E.5.1) Treat Order and Carryover

It is possible to test if the order of treat has an effect on the response. If there is no carryover, no significant effect due to treat order should be observed. A Treat Order variable was created for the two order cases (AB and BA). The test matrix included a TreatOrder variable (coded as 0,1 for BA vs AB), VehType, and miles coded as (0,8,and 16). The model was fit in the usual way with VehType fit as a random variable. The natural log of the FTP (ln(FTP)) was fit as the response. The outlier vehicles (FF4 and GP-1) were not included in this analysis.

The S-plus code was the following:

S-Code: treatoder.lme_lme(log(FTP)~miles+TreatOrder, random=~miles|VehType)

The following ANOVA was produced.

(E.5.1.1) Table: ANOVA to test for Treat Order (Carryover) Effect

	numDF	denDF	F-value	p-value
(Intercept)	1	69	20.64889	<.0001
miles	1	69	9.93901	0.0024
TreatOrder	1	69	6.02426	0.0166

The (TreatOrder) term is highly significant. (p< .02). If there were no carryover, there should be no significant difference as a function of Treat Order. In fact, this general pattern as a function of Treat Order (Carryover) is apparent in each of the vehicle types.

The nature of the carryover becomes more visible if data from each treat order is analyzed separately. The model was fit with vehicle type as the random block, and miles as the fixed effect.

(E.5.1.2) Case 1: Run 1 is Vektron Present, Run 2 is Ref. Fuel (Order AB)

(E.5.1.2.1) Table: ANOVA of Case 1 *(AB Order)*

	numDF	denDF	F-value	p-value
(Intercept)	1	31	19.32623	0.0001
Miles	1	31	1.90861	0.1770

Zero Slope, little increase in NOx for 16,000 miles

The miles term (which tests for a zero slope) is not significant. That is, the slope is depressed or flat across the 16,000 miles. However, modeling data following the other treat order (BA) produces a very different effect.

(E.5.1.3) **Case 2: Run 1 is Ref. Fuel and Run 2 is Vektron Present (*Order BA*)**

(E.5.1.2.1) Table: ANOVA of Case 2 *(BA Order)*

	numDF	DenDF	F-value	p-value
(Intercept)	1	31	26.44300	<.0001
Miles	1	31	12.07494	0.0015

Significant increase in NOx.

The miles term is highly significant in Case 2 indicating a non zero slope. It is tempting to say that this finding is consistent with the fact that alternating and constant fueling groups show the same effect. However, when large carryover is present in a two period crossover design, care needs to be taken when drawing conclusions from an analysis containing Run 2 data.

The literature teaches that because there is aliasing due to carryover, no reliable estimate of treatment effect can be obtained by inclusion of the Run 2 data. Run 1 data is a parallel test in itself and provides an unbiased estimate of the additive effect. Nothing that happened in Run 2 affected the results from Run 1.

(F.0) References

Beckman, R. J., Cook, R.D., (1983), "*Outlier….s"*, Technometrics, Vol 25. No. 2., 119-148.

Bitting, W.H, Firmstone, G.P., and Keller, C.,T. (1995) "A Fleet Test of Two Additive Technologies Comparing Their Effects on Tailpilpe Emissions, Biting," Gasoline Additives Emissions and Performance SAE SP-1095, Feb: 61-84

Burns, V.R., et al. Description of Auto/Oil Quality Improvement Research Program, Society of Automotive Engineers Publication, SAE 912320, 1991

Cotton, J. W. (1989) "Interpreting Data from Two-Period Crossover Design also Termed the Replicated 2x2 Lastin Square Design" Psychological *Bulletin* 106 (3). 503-515.

Fleet Test Team "EPA-OMS, Original Equipment Makers and Infineum Discussions", *Internal Presentation given at EPA Ann Arbor*

Freeman, P.R. (1989) "The Performance of the Two-Stage Analysis of Two-Treatment Two-Period Crossover Trials.", *Statistics in Medicine* 8 (12): 1421-1432

Guerrieri, D.A., Caffrey, P.J., and Rao, V. (EPA) (1995) "Investigation into the Vehicle Exhaust Emissions of High Percentage Ethanol Blends" Gasoline Additives Emissions and Performance SAE SP-1095, Feb :85-103.

Grieve, A.P. (1987) "A Note on the Analysis of the Two-Period Crossover Design when the Period-Treatment Interaction is Significant." *Biomedical Journal* 29 (7).: 771-775

Grizzle, J. E. (1965) "Two-Period Change-Over Design and Its Use in Clinical Trials," *Biometrics*, v. 21: 467

Hecker, H. (1986) "Identification and Interpretation of Effects in Two-Period Crossover Designs." *Edv in Medizin und Biologie* 17 (3). (RECD. 1987): 60-66

Kim, P. J. and Jennrich, R. I. (1973). Tables of the exact sampling distribution of the two sample Kolmogorov-Smirnov criterion. In Selected Tables in Mathematical Statistics, Vol. 1. H. L. Harter and D. B. Owen, eds. Providence, Rhode Island: American Mathematical Society.

Korotney, D.J., Rao, V., Lindhjiem, C.E., and Sklar, M. S. (1995) "Reformulated Gasoline Effects on Exhaust Emissions: Phase III; Investigation of the effects of Sulfur, Olefins, Volatility, and Aromatics and Interactions Between Olefins and Volatility or Sulfur" Gasoline Additives Emissions and Performance SAE SP-1095, Feb: 179-186

Lehmacher, W. (1991) "Analysis of the Crossover Design in the Presence of Residual Effects.", *Statistics in Medicine* 10 (6). : 891-899.

Lindhjem, C.E. (EPA) (1995) "The Effect of Gasoline Reformulation and Sulfur Reduction on Exhaust Emissions from Post-1983 but Pre-1990 Vehicles" Gasoline Additives Emissions and Performance SAE SP-1095, Feb :97-142

Littell, R. C., Milliken, G. A., Stroup, W. W. and Wolfinger, R. D. (1996). SAS System for Mixed Models, SAS Institute, Inc.

Painter, L..J., Rutherford, J.A., Statistical Design and Analysis Methods for the Auto/Oil Air Quality Research Program, Society of Automotive Engineers Publication, SAE 920319, 1992

Pinheiro, J.C., Bates, D.M. (2000) Mixed-Effects Models in S and S-Plus, Springer-Verlag, New York

Putt, M.., Chinchilli, V.M. (1999) "A Mixed Effects Model for the Analysis of Repeated Measure Cross-Over Studies." *Statistics in Medicine* 18 (22): Nov 30: 3037-3058

Ragazzii, R., Nelson, K., (1999*) "*The impact of 10 % Ethonal Blended Fuel on the Exhaust Emissions of Tier 0 and Tier 1 Light Duty Gasoline Vehicles at 35 F." *Mobile Sources Program, Air Pollution Control Division, Colorado Department of Public Health and Environment.*

Vonesh, E.F., Chinchilli, V.M. (1997) Linear and Nonlinear Models for the Analysis of Repeated Measurements, Marcel Dekker, New York

Willian, A.R., Pater J.L. (1986) "Carryover and the Two-Period Crossover Clinical Trial." *Biometrics* 42 (33). : 593-600

Willian, A.R, Pater, J. L. (1986) "Using Baseline Measurements in the Two-Period Crossover Clinical Trial." *Controlled Clin Trials* 7(4).: 282-289

Woods, J. R., Williams, J.G., Tavel, M.. (1989) "The Two-Period Crossover Design in Medical Research." *Annals of Internal Medicine* 110 (7) : 560-565

G.0 Appendix

(G..1) S-plus Output

Model 1a: dlnDiff= ln(Run 1) - ln(SOTB)

```
Linear mixed-effects model fit by REML
 Data: Jun28Dat
        AIC       BIC   logLik
  -14.99587 -10.28366 11.49794

Random effects:
 Formula:  ~ 1 | vehtype
        (Intercept)  Residual
StdDev:   0.1017868 0.1133974

Fixed effects: FirstDiff.1 ~ AdditiveRun1
                  Value  Std.Error DF   t-value p-value
 (Intercept)  0.1438593 0.04984419 18  2.886179  0.0098
AdditiveRun1 -0.0909186 0.04488706 18 -2.025496  0.0579
 Correlation:
             (Intr)
AdditiveRun1 -0.45

Standardized Within-Group Residuals:
       Min         Q1       Med       Q3      Max
 -1.546034 -0.7630491 0.1176449 0.645679 2.002428

Number of Observations: 26
Number of Groups: 7
```

ANOVA

```
                numDF denDF    F-value   p-value
 (Intercept)     1    18       4.888377  0.0402
AdditiveRun1     1    18       4.102633  0.0579
```

Model 1b: dPercDiff = (Run 1- SOTB)/SOTB

```
Linear mixed-effects model fit by REML
 Data: Jun28Dat
        AIC       BIC  logLik
  -9.325839 -4.613624 8.66292

Random effects:
 Formula:  ~ 1 | vehtype
        (Intercept)  Residual
StdDev:   0.1150728 0.1274823

Fixed effects: PercDiff1 ~ AdditiveRun1
                  Value  Std.Error DF   t-value p-value
 (Intercept)  0.1683203 0.05622366 18  2.993763  0.0078
AdditiveRun1 -0.1042360 0.05046390 18 -2.065556  0.0536
 Correlation:
             (Intr)
AdditiveRun1 -0.449

Standardized Within-Group Residuals:
       Min         Q1        Med         Q3      Max
 -1.422782 -0.7722425 0.07673865 0.6650237 2.087697

Number of Observations: 26
Number of Groups: 7
```

ANOVA

	numDF	denDF	F-value	p-value
(Intercept)	1	18	5.348881	0.0328
AdditiveRun1	1	18	4.266521	0.0536

(G.2) SAS Output

(G.2.1) Response: <u>PercDiff</u>

```
Analysis of Run 1 data for the EPA Field Test (minus GP1 and FF4)  13
12:31 Monday, January 22, 2001

                                      The Mixed Procedure

                                       Model Information

                     Data Set                     EPA.RUN1CALC
                     Dependent Variable           PERCDIFF
                     Covariance Structure         Variance Components
                     Estimation Method            REML
                     Residual Variance Method     Profile
                     Fixed Effects SE Method      Model-Based
                     Degrees of Freedom Method    Containment


                                    Class Level Information

                      Class      Levels    Values

                      VEHTYPE         7    DC EX FE FF GC GP HA


                                          Dimensions

                           Covariance Parameters              2
                           Columns in X                       2
                           Columns in Z                       7
                           Subjects                           1
                           Max Obs Per Subject               26
                           Observations Used                 26
                           Observations Not Used              0
                           Total Observations                26
```

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	-12.31330484	
1	2	-17.32581172	0.00000089
2	1	-17.32583917	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate
VEHTYPE	0.01324
Residual	0.01625

The Mixed Procedure

Fit Statistics

Res Log Likelihood	8.7
Akaike's Information Criterion	6.7
Schwarz's Bayesian Criterion	6.7
-2 Res Log Likelihood	-17.3

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	0.1683	0.05622	6	2.99	0.0242
ADDITIVE	**-0.1042**	**0.05046**	**18**	**-2.07**	**0.0536**

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
ADDITIVE	1	18	4.27	0.0536

(G.2.2) Response: dlnDiff

The Mixed Procedure

Model Information

Data Set	EPA.RUN1CALC
Dependent Variable	LNDIFF
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
VEHTYPE	7	DC EX FE FF GC GP HA

Dimensions

Covariance Parameters	2
Columns in X	2
Columns in Z	7
Subjects	1
Max Obs Per Subject	26
Observations Used	26
Observations Not Used	0
Total Observations	26

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	-18.05252099	
1	2	-22.99584355	0.00000089
2	1	-22.99587349	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate
VEHTYPE	0.01036
Residual	0.01286

The Mixed Procedure

Fit Statistics

Res Log Likelihood	11.5
Akaike's Information Criterion	9.5
Schwarz's Bayesian Criterion	9.6
-2 Res Log Likelihood	-23.0

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	0.1439	0.04984	6	2.89	0.0278
ADDITIVE	**-0.09092**	**0.04489**	**18**	**-2.03**	**0.0579**

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
ADDITIVE	1	18	4.10	0.0579

G.3 Likelihood Ratio Tests for FSG Term for Models Fit Using (ML) rather than (REML)

(These are similar results to those observed in Section E.3)

(G.3.1) Likelihood Ratio Test for FSG Term for Response: dlnDiff

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
mod1a.ml	1	5	-24.49809	-18.20761	17.24905			
mod2a.ml	2	4	-23.85935	-18.82697	15.92968	1 vs 2	2.638739	0.1043

(G.3.2) Likelihood Ratio Test for FSG Term for Response: PercDiff

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
mod1b.ml	1	5	-18.41545	-12.12497	14.20772			
mod2b.ml	2	4	-17.71251	-12.68013	12.85626	1 vs 2	2.702937	0.1002